



Turkish and Native English Academic Writers' Use of Lexical Bundles*

Yusuf Öztürk ^a **, Gül Durmuşoğlu Köse ^b

^a *Anadolu University, Eskişehir, Turkey*

^b *Anadolu University, Eskişehir, Turkey*

APA Citation:

Öztürk, Y., & Köse, G. D. (2016). Turkish and native english academic writers' use of lexical bundles. *Journal of Language and Linguistic Studies, 12(1), 149-165.*

Abstract

Lexical bundles such as *on the other hand* and *as a result of* are extremely common and important in academic discourse. The appropriate use of lexical bundles typical of a specific academic discipline is important for writers and the absence of such bundles may not sound fluent and native-like. Recent studies (e.g. Adel & Erman, 2012; Chen & Baker, 2010) have revealed that non-native writers produce not only fewer types of lexical bundles, but also less varied ones. Furthermore, they also overuse a restricted number of bundles in their writing. Focusing on this issue, this study aimed to investigate Turkish and native English postgraduate students' and native scholars' use of lexical bundles in a specific academic discipline, that is foreign language teaching, in terms of frequency, functions and structures. For this aim, a corpus of 150 texts was collected containing Turkish and native English students' MA and PhD theses along with native scholars' published research articles. Four-word lexical bundles were identified using WordSmith Tools 6. The results revealed that Turkish postgraduate students used far more lexical bundles in their texts compared to both native students and scholars. However, there was a redundancy in Turkish students' texts when the token frequencies were examined, meaning that Turkish students overused most of the lexical bundles. On the other hand, statistical analysis of the bundle lists revealed that Turkish postgraduate students employed different bundles from their native peers and scholars. Finally, the structural and functional categories of the lexical bundles did not show any statistically significant differences across the research sub-corpora.

© 2015 JLLS and the Authors - Published by JLLS.

Keywords: Lexical bundles, English academic discourse, Turkish students, native English students, corpus linguistics

1. Introduction

In recent decades, English has become the lingua franca in the academia and a global means of communication for the dissemination of knowledge and science (Björkman, 2013). Students and scholars are thus expected to show a native-like proficiency in this global language to be able to carry out research and publish their works. They also need to be familiar with "the distinguishing features of academic discourse such as vocabulary, norms, set of conventions, and modes of inquiry" (Zamel, 1998, p. 187). With the advent of technology in the last decades, it has been made possible to examine

*This paper is based on a Master's thesis submitted to Anadolu University Graduate School of Educational Sciences and funded by Anadolu University Scientific Research Projects Commission under the grant no. 1305E093. A preliminary version was presented at the AAAL 2014 conference held in Portland, Oregon, USA on March 22-25, 2014.

** Corresponding author. Tel.: 0506 299 98 94
E-mail address: yusufozturk@anadolu.edu.tr

naturally occurring lengthy texts, and reveal patterns in language use and identify distinguishing features of a register. In this respect, there have been many studies which particularly focused on academic writing and revealed that “language in use is characterized by repetition of fixed and semi-fixed multi-word combinations and by use of formulaic patterns” (Byrd & Coxhead, 2010, p. 32). Such studies examined multi-word combinations in both non-native academic writing (e.g. Wei & Lei, 2011; Hyland, 2008a-b) and general academic writing (e.g. Liu, 2012; Byrd & Coxhead, 2010). One type of these multi-word combinations that have been thoroughly studied in the literature is “lexical bundles”, which refers to expressions of three or more words that frequently co-occur in a corpus. They are extremely common and important in academic discourse, and are argued to be an important component of fluent linguistic production and a crucial part of native-like proficiency (Cowie, 1998; Hyland, 2012; Simpson-Vlach & Ellis, 2010). Although it was reported in the previous literature that non-native writers would produce fewer multi-word expressions overall (Erman, 2009; Howarth, 1998) and less varied ones (Granger, 1998; Lewis, 2009) than native writers, several studies revealed conflicting results (Wei & Lei, 2011; Hyland, 2008b). On the other hand, the literature on the lexical bundle use of Turkish speakers of English has been extremely limited. In this respect, this study focuses on Turkish and native English postgraduate students’ and native scholars’ use of lexical bundles in academic writing in order to see to what extent Turkish students’ use of lexical bundles can approximate to that of their native peers and scholars. In addition to contributing to the literature in Turkey, such a study would also enhance our general understanding of the use of lexical bundles in academic writing.

1.1. Literature review

1.1.1. Definition of lexical bundles

The term ‘lexical bundles’, firstly used by Biber et al. (1999), can be briefly described as expressions of three or more words that show a statistical tendency to co-occur in a particular corpus and identified based on a standardized frequency and distribution criteria. Examples include I don’t know what or I said to him in conversation, and as a result of or on the other hand in academic prose. What is remarkable about lexical bundles is that they are extremely common and constitute an important part of discourse. Biber et. al. (1999) found that 21% of all the words in their academic prose corpus occurred in a recurrent lexical bundle. Beside their recurrent nature, lexical bundles also have particular characteristics distinguishing them from other types of multi-word expressions like collocations and idioms. One of these characteristics is that “most lexical bundles are not idiomatic in meaning and not perceptually salient” (Biber & Barbieri, 2007, p. 269). In other words, the meaning of a lexical bundle only by looking at its individual items can easily be understood, unlike idioms where more than the literal meaning of the items is needed. Another characteristic is that lexical bundles are not usually complete structural units as in the examples of in the case of and the base of the (Biber & Barbieri, 2007), but they are mostly part of longer structures. Furthermore, lexical bundles, as seen in the examples, include both function words and content words, as opposed to collocations which usually consist of content words.

Lexical bundles are extremely common in language use as mentioned above, but what makes them even more important for people writing for academic purposes is that they vary across different disciplines (Hyland, 2012). This means that appropriate use of lexical bundles typical of a specific academic discipline is important for writers, and the absence of such bundles may reveal “the lack of fluency of a novice” (p. 165). There is no doubt that another dimension of difficulty is also added for the writers who are the non-native speakers of the language they are writing in (Adel & Erman, 2012). In the context of Turkey, for example, Turkish academics and postgraduate students are usually

required to publish their research in English so that they can fulfil their studies and progress in their academic career as well as contribute to the international literature. To do this, they need to have a certain level of English proficiency and also be familiar with the common lexical bundles used in their disciplines so that they do not sound as novice writers.

1.1.2. Studies on lexical bundles

Although lexical bundles are reported to be very frequent in academic prose and a component of fluent linguistic production, different studies on non-native EFL/ESL speakers' use of lexical bundles (e.g. Adel & Erman, 2012; De Cock, Granger, Leech & McEnergy, 1998) showed that non-native writers produce not only fewer types of lexical bundles, but also less varied ones, compared to native English writers. Similarly, some studies also found that non-native writers overuse a restricted number of bundles (De Cock et al., 1998; Wei & Lei, 2011).

As one of the early studies, De Cock et al. (1998) used the term formulaic expressions referring to automatically extracted combinations of two, three, four and five words, and examined the formulaic competence of advanced adult EFL learners of French L1 in a corpus of informal speech. Though their data set (i.e. informal speech) is quite different from the focus of this study (i.e. academic prose), their results were important as being one of the first studies in the literature. They found that advanced EFL learners made use of multi-word combinations, and in some cases, even more combinations than native speakers. However, they reported that the learners' use was 'not necessarily the same as those used by the native speakers' in terms of frequency, syntactic uses and pragmatic functions (p. 78). Since then, a number of corpus-based studies investigated the use of lexical bundles from a variety of perspectives including variations between different registers, disciplines and groups of writers with different L1s and writing expertise.

Examining a 3.5-million-word corpus containing 120 published papers in four disciplines (30 papers in each), and 80 PhD and Master's theses (20 in each disciplines) of students at five Hong Kong universities, Hyland (2008) compared the use of lexical bundles in the texts by different levels of writers. He found that the frequency of forms, structures and functions varied considerably across student and expert writing. He reported that the research articles contained fewer lexical bundles and fewer different lexical bundles overall, and included largely different lexical bundles compared to the PhD and Master's theses. In a study with a similar corpus, Wei and Lei (2011) investigated the use of lexical bundles in a corpus of doctoral dissertations by Chinese L1 learners and published journal articles by professional writers. Supporting Hyland (2008), the findings showed that the advanced learner writers used much more bundles and different bundles than the professional writers did. Two recent studies (Chen & Baker, 2010; Adel & Erman, 2012) examined the writings of university students. Chen and Baker (2010) found that the native English student and expert writing contained more types of lexical bundles than the Chinese students' did. They also argued that non-native writers had some control of these bundles, but do not "demonstrate it as diversely and robustly as native writers do" (p. 43). Focusing on the essays of Swedish and British university students written in a specific discipline, that is linguistics, Adel and Erman (2012) found that native students' texts contained a far wider range of bundles than those of non-native students. Moreover, frequency of the 70% of the bundles used by one group (43 types in non-native data and 89 in native data) differed statistically significantly from the other.

1.1.3. Studies on the lexical bundle use of Turkish writers

To our knowledge, only two studies in the literature examined the use of lexical bundles by Turkish L2 writers. Bal (2010) investigated the use of lexical bundles in research articles written in English by Turkish scholars, and reported the most frequent lexical bundles as *on the other hand*, *the end of the*, *as well as the*, *in the case of* and *one of the most*, out of the 99 bundles identified at 20 times per million words. In other words, she merely described the lexical bundles used by Turkish scholars, and did not examine how their use approximate native speakers of English. On the other hand, Karabacak and Qin (2012) investigated the use of lexical bundles in argumentative papers written by three groups of university writers, Turkish, Chinese, and Americans. Their analysis revealed that 96 bundles were used by Turkish and Chinese students but never used by American students. And they concluded that some bundles are not acquired naturally, meaning that simple exposure does not transfer directly into students' production in writing. Therefore, they suggested that explicit teaching might be required to hasten their acquisition process. However, their study did not include an in-depth analysis into the structures and functions of lexical bundles although Biber, Conrad and Reppen (1998) argue that it should, and also used a relatively small research corpus.

1.2. Research questions

To contribute to the limited research on the use of lexical bundles by Turkish students, and shed some light to the conflicting findings reported in the literature, this study aimed to investigate Turkish and native English postgraduate students' and native scholars' use of lexical bundles in a specific academic discipline, that is foreign language teaching, in terms of frequency, functions and structures of bundles. In this regard, the following research questions were addressed in the study:

- (1) Which lexical bundles are frequently used by Turkish and native English postgraduate students and native scholars?
- (2) To what extent do Turkish and native English postgraduate students and native scholars differ in terms of:
 - (a) type and token frequency of the lexical bundles,
 - (b) their structures,
 - (c) and functions?

2. Method

2.1. Research corpus

This study used a small and specialised corpus based on the aims of the study. Although Sinclair (2004) asserts that "small is not beautiful" (p. 189) when it comes to building a corpus, small corpora better suit the teaching contexts with specific needs such as ESP or EAP (Flowerdew 2002; Tribble, 2002). Furthermore, while large corpora provide insights into the patterns in the language as a whole, small and specialized corpora "give insights into patterns of language use in particular settings" (Koester, 2010, p. 67). In this regard, a research corpus with three main sub-corpora was compiled, and it included Turkish and native English postgraduate students' MA/PhD theses, and native scholars' published research articles as baseline. These genres were chosen as they "represent the key research genres of the academy" (Hyland, 2008a, p. 47). In addition, research articles written by native English scholars were included because native peer writing does not always include ideal and standard usage, and research articles can thus provide useful data when combined with student writing (Chen, 2009). The reason why an equivalent sub-corpus of research articles by Turkish scholars was not

compiled in the study was that the primary focus of this study was the academic writing of postgraduate students, and as just mentioned, native speaker research articles were only included because they were considered to represent baseline data to make a comparison with postgraduate student writing. Table 1 represents the research corpus and its sub-corpora along with the number of texts and words they contained.

Table 1. Distribution of the total number of words and texts

		No. of Texts	No. of Words	Total
TPMPT	MA	30	612.379	1.346.396
	PhD	20	734.017	
NPMPT	MA	30	457.594	1.239.392
	PhD	20	781.798	
NSA		50	446.009	446.009
	Total	150		3.031.797

NSA: Native Scholars' Articles NPMPT: Native Postgraduate Students' MA/PhD Theses

TPMPT: Turkish Postgraduate Students' MA/PhD Theses

2.2. Identification of lexical bundles

The present study focused on four-word lexical bundles for two reasons. Firstly, four-word bundles are the most studied length in such studies and considered to be manageable in size for further analysis (Chen & Baker, 2010). Secondly, they are “over 10 times more frequent than five-word sequences and offer a wider variety of structures and functions to analyze” (Hyland, 2012, p. 151). Another issue that is of significance for identifying lexical bundles in a corpus is the frequency cut-off point. Cut-off points used in the literature vary from 10 times (Biber et. al., 1999) to 40 times (Biber & Barbieri, 2007) per million words. They are regarded as “somewhat arbitrary” (Biber & Barbieri, 2007, p. 267.), and are usually decided based on the size of the corpus. Considering the size of the corpus used in this study, 25 times per million words was set to be the frequency cut-off point. Finally, a distribution criterion, which is occurring in at least five texts in each sub-corpus, was also adopted to avoid individual idiosyncrasies (Biber, Conrad & Cortes, 2004).

The identification process was performed by means of WordSmith Tools 6 (Scott, 2011). Before computing the texts, all the direct quotations were deleted since the writers' own use of lexical bundles was the focus. In addition, all the tables/figures, end/foot notes, and references/appendices were excluded, leaving back only plain text produced by the writers. For the first research question, all four-word combinations occurring at least 25 times per million words and in five texts were retrieved automatically. Then, content/context-dependent bundles such as *second language acquisition process* or *in the Turkish context* were excluded since they needed “to be removed as they are not the ‘building blocks’ which carry a distinct discourse function” (Chen, 2009, p. 58) and overlapping bundles such as *it has been suggested* and *has been suggested that* were combined into a five-word bundle as in *it has been suggested that* to avoid inflated results (Chen & Baker, 2010). For the second research question, it was examined to what extent the three sub-corpora differed based on the type/token frequency, function and structure of the lexical bundles retrieved. In addition to comparing raw type/token frequencies, a log-likelihood analysis was also done by using the KeyWords function of WordSmith to see whether there was a statistically significant difference across the three sub-corpora in terms of the frequencies of the lexical bundles.

2.3. Structural categorisation

The lexical bundles identified in the study were structurally categorised and compared across the three sub-corpora. For this categorisation, Biber et al.'s taxonomy (1999) was used, as it is the only

taxonomy encountered in the literature with slight adaptations. It includes twelve structural categories such as noun phrase with *of*-phrase (*the end of the*), anticipatory *it* + verb (*it is possible to*) and passive verb + prepositional phrase fragment (*are shown in table*). Chi-square test was done to see whether there are significant differences between Turkish and native English postgraduate students and scholars in terms of structures of the lexical bundles.

2.4. Functional categorisation

Final step of the analysis included the functional categorization of the lexical bundles identified in the corpus. With regard to this, the widely used taxonomy initially designed by Cortes (2002), and later improved in Biber et al. (2004 & 2007) was used in this study. The taxonomy includes three primary discourse functions, which are (1) stance expressions, (2) discourse organizers, and (3) referential expressions (Biber and Barbieri, 2007, pp. 270). Stance bundles such as *are more likely to* and *it is important to* are used to express attitudes or assessments in terms of certainty or uncertainty that frame some other proposition. Discourse organizers such as *on the other hand* and *in contrast to the* express the connections between prior and coming discourse. On the other hand, referential bundles including *at the beginning of* and *in the current study* make direct reference to physical or abstract entities, or to the textual context itself, either to identify the entity or to single out some particular attribute of the entity as especially important. The same as the structural analysis, chi-square test was conducted right after the lexical bundles identified in the corpus were categorized functionally to reveal whether any significant differences exist between Turkish and native texts.

3. Results and Discussion

3.1. Overall results

After the content/context dependent bundles were excluded and the overlapping bundles were combined, the whole research corpus contained a total of 271 lexical bundles. Among these, 125 lexical bundles were identified in the Turkish students' MA and PhD theses, 77 lexical bundles in native scholars' research articles, and 69 lexical bundles in native English students' theses. In other words, the number of lexical bundles in the Turkish students' texts was almost as twice as those in the native scholars and native students. With regard to the number of bundle types, the native English students and scholars showed a similar pattern, but the Turkish students were found to use a far wider range of different lexical bundles in their texts, which can be interpreted as using many different lexical bundles quite repetitively in their writing. These findings are consistent with those of Hyland (2008b) and Wei and Lei (2011). This can be because they also focused on postgraduate theses. In these studies, Chinese and Cantonese L1 students' theses included much more lexical bundles than research articles which might have been written by a native or a non-native speaker of English. What is also common among these studies is the repetitive nature of the non-native texts, which was also revealed in the present study. As a result, since these studies also focused on advanced academic writing (i.e. theses/articles) as the current study, it can be inferred that when it comes to advanced academic writing, non-native writers including Turkish L1 writers tend to employ considerably higher number of bundle types in a much more repetitive way, differing from native English writers. This argument can be supported by referring to the findings of Chen and Baker (2010) and Adel and Erman (2012) that focused on argumentative essays by undergraduate students: they revealed that Swedish and Chinese students employed lower number of bundles than native speakers.

With respect to the actual lexical bundles most frequently used by the three groups of writers, the findings are presented in Table 2 below. It presents the 50 most frequent bundles in order of token

frequency, and bundles shared by the three groups are shaded in gray while those bundles of Turkish writers' shared by one of the other two groups emphasized in brown.

Table 2. List of the 50 most frequent lexical bundles identified in the research corpus

NSA	#	NPMPT	#	TPMPT	#
in the current study	83	(at) + the end of the	154	at the end of + (the)	567
in the present study	65	it is important to	153	on the other hand	503
the extent to which	61	at the same time	151	the results of the	357
the results of the	54	as well as the	150	(at) + the beginning	325
on the other hand	47	on the other hand	139	as a result of + (the)	324
in the case of	46	the results of the	125	end of the study	310
(at) + the end of the	42	as a result of	105	beginning of the	198
it is important to	39	at the beginning of	92	the analysis of the	182
on the basis of	38	in the present study	82	of the present study	177
the nature of the	37	in the form of	77	in terms of the	166
it is possible that	36	the results of this +	74	in the present study	155
for each of the	35	the use of the	72	with the help of	129
at the same time	35	the total number of	68	at the same time	129
in the context of	32	to be able to	66	the findings of the +	122
the results of this +	30	the purpose of this	62	in the light of	119
in the form of	28	through the use of	61	to be able to	110
of the current study	28	to the fact that	59	one of the most	109
as well as the	27	in addition to the	58	in the use of	104
it is clear that	25	used in this study	57	to find out the	102
as a function of	25	in terms of the	57	that there is a	101
of the present study	24	in a variety of	54	is one of the	92
the total number of	24	the rest of the	54	as can be seen + (in)	87
with respect to the	24	in the current study	54	as well as the	86
the fact that the	22	in other words the	53	results of the study	83
were more likely to	22	in the case of	53	is considered to be	83
over the course of	21	for the purpose of	50	in addition to the	83
as a result of	20	is important to note	50	on the use of	82
in addition to the	20	in the following	49	by the help of	82
with the exception of	20	at the time of	48	in order to find	80
the effect of the	20	the fact that the	48	in order to see	79
to ensure that the	19	a great deal of	48	in the field of	78
are presented in table	19	of the present study	47	the fact that the	77
in a way that	18	in the next section	47	the aim of the	77
the degree to which	18	the majority of the	45	to find out whether	76
in contrast to the	17	the role of the	45	in the form of	74
in the same way	17	in the context of	44	it can be concluded	73
at the time of	17	on the part of +	44	the results of this	69
used in this study	17	the way in which	44	it was found that	69
a number of studies	17	can be found in	43	in other words the	69
in relation to the	17	in an attempt to	42	that most of the	68
there was also a	17	in a way that	41	the purpose of the	66
at the beginning of +	16	for the purposes of	41	it can be said +	66
that there is a	16	as well as a	40	that there was a	63
it should be noted +	16	one of the most	40	in line with the	63
in terms of the	16	as a result the	40	that the use of +	61
(as) + can be seen in	16	for each of the	39	of the fact that	61
the purpose of this	16	I was able to	39	in addition to this	59
the purpose of the	15	in an effort to	38	according to the	59
in the field of	15	has been shown to	38	to the fact that	58
to the fact that	15	due to the fact	38	the findings of this	56

NSA: Native Scholars' Articles NPMPT: Native Postgraduate Students' MA/PhD Theses

TPMPT: Turkish Postgraduate Students' MA/PhD Theses

As is seen in the table, the most frequently used lexical bundle in Turkish writers' theses was *at the end of + (the)*, which was used 567 times and also the most frequent bundle in native English students' theses with a frequency of 154 times. As for the native scholars' published research articles, the most frequent bundle was in the current study with a frequency of 83, although it was not among the 50 most frequent bundles in the theses. Examining the table above, it can be easily seen that almost half of the 50 most frequent bundles in Turkish writers' theses were also used in native writers' theses and/or published research articles. Furthermore, many of the other bundles are actually variants of the shared bundles. To give an example, in addition to *the* was shared by the three groups of writers, but Turkish writers also used *in addition to this* which was not preferred by native English writers. Similarly, *end of the study* also appeared in Turkish writers' theses in addition to *(at) + the end of*, but not in those of native writers.

Despite the huge difference in the number of bundle types, based on the most frequently used 50 bundles, Turkish writers seem to employ similar lexical bundles with those of their native peers and native scholars. However, there were some bundles employed by Turkish writers, but never or very rarely occurred in native English writers' texts, and vice-versa. For instance, *with the help of* and *by the help of* are among those bundles. *By the help of* never occurred in native writers' theses and research articles while *with the help of* had a frequency of 12 times in total in opposed to 129 times in Turkish writers' theses. On the other hand, native English writers preferred *through the use of*, probably to denote a similar notion with *with the help of* and *by the help of*. Another example of this can be *is considered to be* which occurred 83 times in Turkish writers' theses, but only 14 times in native writers' theses and research articles together. Instead of *is considered to be*, native English postgraduate students and scholars preferred different and usually more powerful stance bundles such as *it is important to*, *it is possible that* and *were more likely to* which Turkish students very rarely used:

Although Turkish writers seem to employ similar bundles with native English writers especially when it comes to frequently used bundles, the fact that Turkish writers' texts have a quite repetitive nature still stands. As an example, *on the other hand* was used 47 times by native established scholars and 109 times by native writers. However, Turkish writers employed *on the other hand* 503 times, almost 10 times more than native established scholars and 5 times more than native writers. The case of *at the end of + (the)* is also the same. It occurred 567 times in Turkish theses, 154 times in native theses, and 42 times in the research articles.

3.2. Statistical significance

To see whether a bundle in a sub-corpus is statistically significantly overused or underused with reference to another sub-corpus, KeyWord function of WordSmith was used. Firstly, Turkish and native English postgraduate students' bundles were compared with reference to those of native English scholars (see Table 3 below). Secondly, native postgraduate students' and scholars' bundles were then compared with reference to Turkish postgraduate students (see Table 5). In the tables, the lexical bundles that were not shared by Turkish and native writers and statistically significantly differed in frequency were shaded in bold.

Table 3. Key lexical bundles in TPMPT and NPMPT with NSA as the reference corpus
($p < .001$)

Corpus		Key lexical bundles
TPMPT	(+)	end of the study (177,43), at the end of + (the) (151,95), beginning of the study (113,32), (at) + the beginning of the (108,31), on the other hand (97,55), as a result of + (the) (88,46), with the help of (73,83), the findings of the + (study) (69,82), in the light of (68,11) , to be able to (62,95), one of the most (62,38), to find out the (58,38) , is one of the (52,65) , is considered to be (47,50) , results of the study (47,50), on the use of (46,93) , by the help of (46,93) , in order to find (45,78) , in order to see (45,21) , the analysis of the (44,80), the aim of the (44,07) , to find out whether (43,50) , it can be concluded + (that) (41,78) , in other words the (39,49), it was found that (39,49) , that most of the (38,92) , it can be said + (that) (37,77) , that there was a (36,06), in line with the (36,06) , of the fact that (34,91) , that the use of + (the) (34,91) , the results of the (34,77), according to the results (33,77) , in addition to this (33,77) , the findings of this (32,05) , the number of the (31,48) , in terms of the (31,16), it is seen that (30,33) , findings of this study (29,76) , it was seen that (29,19) , it can be claimed (29,19)
	(-)	---
NPMPT	(+)	the use of the (44,26), to be able to (40,58), through the use of (37,50), the rest of the (33,20), in a variety of (33,20), in other words the (32,58), for the purpose of (30,74), in the following example (30,12), a great deal of (29,51), in the next section (28,89)
	(-)	in the current study (-70,15)
	(+): Overuse, (-): Underuse, NPMPT: Native Postgraduate Students' MA/PhD Theses	TPMPT: Turkish Postgraduate Students' MA/PhD Theses

When native scholars' articles taken as reference, 41 bundles in Turkish postgraduate students' theses were statistically significantly overused while only 10 bundles were overused and 1 bundle was underused in native postgraduate students' theses. Again, it could be argued that native postgraduates' use of lexical bundles were closer to that of native scholars, compared to Turkish postgraduate students. The repetitive pattern in Turkish students' texts can be observed here as well; the keyness scores (indicated in parentheses) are much higher in Turkish students' bundles. As emphasized in bold, there are 27 bundles that were not shared by neither native postgraduate students and native scholars, and overused by Turkish students. Although some of these can be regarded as variants of similar bundles that were already shared such as *in addition to this* (shared bundle: *in addition to the*) and *of the fact that* (shared bundle: *the fact that the*), these bundles seem to be unique to Turkish postgraduate students, and clearly not employed by their native peers and native scholars. The similar studies in the literature were reviewed whether these 27 bundles were used by students with different L1 backgrounds. 11 of these bundles were indeed reported to be used by Cantonese (Hyland, 2008), Chinese (Chen & Baker, 2010; Wei & Lei, 2011), Swedish (Adel & Erman, 2012) and Turkish (Bal, 2010) students/writers. The remaining 16 bundles seem to be used only by Turkish postgraduate students based on the literature. Because they were not used by individuals with different L1 backgrounds in the literature, it can be argued that the bundles that seem to be used only in the texts produced by the Turkish postgraduate students may be a transfer from Turkish. In this regard, examining a list of frequently used academic verbs compiled by Yıldız and Aksan (2013) in a one-million corpus of Turkish academic texts in 15 disciplines can be useful. Table 4 presents the 10 most frequently used verbs in academic Turkish.

Table 4. 10 most frequently used verbs in academic Turkish (Yıldız & Aksan, 2013)

Verb	Frequency	Translation
görülmektedir	813	<i>It is seen</i>
göstermektedir	655	<i>It shows</i>
bulunmuştur	541	<i>It was found</i>
gerekmektedir	475	<i>It should...</i>
bulunmaktadır	431	<i>It is found</i>
görülmüştür	403	<i>It was seen</i>
belirlenmiştir	369	<i>It was identified</i>
saptanmıştır	334	<i>It was determined</i>
söylenebilir	296	<i>It can be said</i>
gerekir	292	<i>It should...</i>

Four of the top 10 most frequent academic verbs and their English translations were shaded with bold since they had been identified as being unique to the Turkish postgraduate students in the current study. Based on the table, it may be claimed that Turkish postgraduate students transferred some Turkish expressions to various lexical bundles in English, and consequently, differed from native English postgraduates and scholars.

As for the second significance analysis, the key bundles in native postgraduate students' and scholars' texts were determined with reference to Turkish students' texts. In other words, this analysis reveals the bundles statistically significantly overused or underused in native texts when compared to Turkish texts. The findings are summarized in Table 15. The bundles significantly overused or underused in both native postgraduate students' and scholars' texts with reference to Turkish students' texts were shaded in bold.

Table5. Key lexical bundles in NSA and NPMPT with TPMPT as the reference corpus
(p < .001)

Corpus		Key lexical bundles
NPMPT	(+)	the total number of (100,01), in a variety of (79,42), in the current study (79,42), in the case of (77,95), is important to note + (that) (73,54), in the following example (72,07), it is important to (71,53), a great deal of (70,60), at the time of (70,60), in the next section (69,13), the majority of the (66,18), on the part of + (the) (64,71), the purpose of this + (study) (64,71), the way in which (64,71), can be found in (63,24), in an attempt to (61,77), in a way that (60,30), for the purposes of (60,30), as well as a (58,83), as a result the (58,83), for each of the (57,36), I was able to (57,36), has been shown to (55,89), in an effort to (55,89), are more likely to (54,42), as part of the (51,48), the course of the (50,01), the ways in which (50,01), the context of the (48,53), in order to determine (48,53), it is possible that (48,53), by the end of (45,59), as a way to (45,59) in terms of the (-47,00), of the present study (-70,02), the results of the (-98,09), as a result of (-99,87), the analysis of the (-101,75), at the beginning of + (the) (-108,28), on the other hand (-190,19), (at) + the end of the (-207,23)
	(-)	in the current study (230,90), the extent to which (169,70), in the case of (127,97), it is possible that (100,15), for each of the (97,37), of the current study (77,89), as a function of (69,55), it is clear that (69,55), with respect to the (66,76), the total number of (66,76), were more likely to (61,20), over the course of (58,42), with the exception of (55,64), the effect of the (55,64), to ensure that the (52,86), are presented in table (52,86), the degree to which (50,07), in a way that (50,07), in contrast to the (47,29), at the time of (47,29), a number of studies (47,29), there was also a (47,29), are summarized in table (41,73), the
NSA	(+)	in the current study (230,90), the extent to which (169,70), in the case of (127,97), it is possible that (100,15), for each of the (97,37), of the current study (77,89), as a function of (69,55), it is clear that (69,55), with respect to the (66,76), the total number of (66,76), were more likely to (61,20), over the course of (58,42), with the exception of (55,64), the effect of the (55,64), to ensure that the (52,86), are presented in table (52,86), the degree to which (50,07), in a way that (50,07), in contrast to the (47,29), at the time of (47,29), a number of studies (47,29), there was also a (47,29), are summarized in table (41,73), the
	(-)	

context of the (41,73), these results suggest that (38,95), (is) + important to note that (38,95), should be noted that (38,95), was found to be (36,16), a greater number of (36,16), in the absence of (36,16), are more likely to (36,16), the ways in which (36,16), the focus of the (36,16), to the extent that (36,16), beyond the scope of (33,38), it is likely that (33,38), play a role in (33,38), that the number of (33,38), the size of the (30,60), to be related to (30,60), in an attempt to (30,60), in any of the (30,60), from the current study (30,60), in a study of (30,60), it may be that (30,60), it is difficult to (30,60)

(-) in terms of the (-31,16), the results of the (-34,77), the analysis of the (-44,80), as a result of (-88,46), at the beginning of + (the) (-94,75), on the other hand (-97,55), (at) + the end of the (-129,71)

(+): Overuse, (-): Underuse, NSA: Native Scholars' Articles, NPMPT: Native Postgraduate Students' MA/PhD Theses

Native postgraduate students overused a total of 33 bundles and underused 8 bundles; as for native scholars, they overused 46 bundles, and underused 7 bundles, when compared to Turkish students. As mentioned in the overall findings, the number of bundle types in native scholars' articles was 83, and native postgraduate students 75. Therefore, considering the number of key bundles, it can be argued that Turkish postgraduate students considerably differed from native postgraduate students in their use of lexical bundles. Although they seem to have shared bundles with their native counterparts and scholars, which may show their high level of English and familiarity with academic writing, even the raw frequencies of these bundles differ to a large extent, which points to the verbose or redundant nature in their writing. Furthermore, the bundles unique to Turkish students and not even employed by other L1 writers in similar studies such as *it can be said that* or *it was seen that* could be due to their effort to directly translate what they have in mind in Turkish to English rather than trying to be native-like and more academic. On the other hand, the bundles observed to be unique to native texts in the current study such as *in an attempt to*, *it is clear that* and *was found to be* seem to reveal the bundles distinguishing them from Turkish texts.

3.3. Structures of lexical bundles

The findings regarding the structural categories of the lexical bundles employed by the three groups of writers are represented in Figure 1. The distribution of bundle structures seems to be similar with only minor differences. In this sense, the chi-square test did not reveal any statistically significant differences between the three groups of writers and 12 structural categories, $X^2(22, N = 291) = 23.75$, $p = .36$.

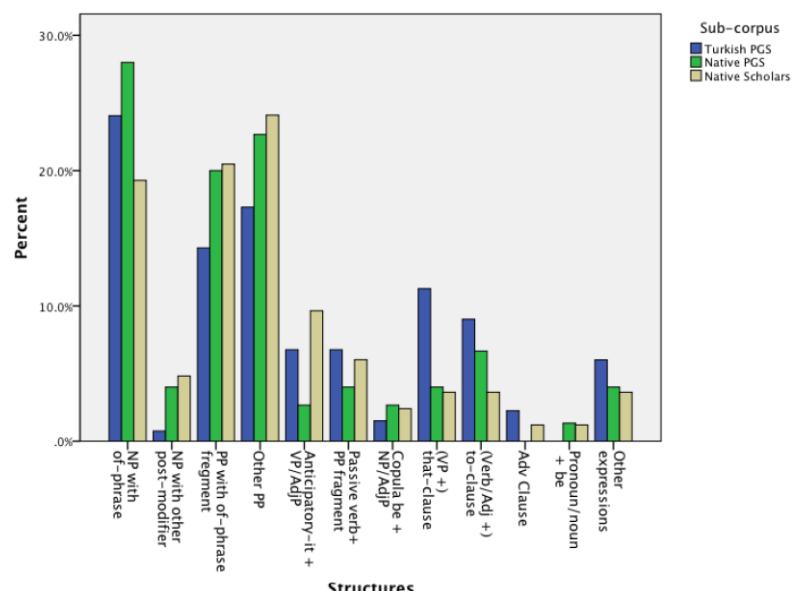


Figure 1. Structural distribution of bundles used by three groups of writers

For NP with of-phrase fragment and NP with other post-modifier fragments, native texts included slightly more NP-based bundles (e.g. *the results of the, the extent to which*). Likewise, PP-based bundles (e.g. *at the end of, with respect to the*) also occurred more in native texts than that of Turkish texts. As for VP-based bundles, anticipatory-it + VP/AdjP (e.g. *it is important to, it can be concluded*) and passive verb + PP (e.g. *can be seen in, are summarized in table*) fragment structures were used more in Turkish and native scholar texts than native student texts. Copula be + NP/AdjP bundles (e.g. *is important to note, is one of the*) were distributed almost equally and formed a very small proportion. Two types of structures that Turkish students employed more frequently were (Verb phrase +) that-clause fragment (e.g. *the results showed that, we can say that*) and (Verb/Adjective +) to-clause fragment (*to be able to, are more likely to*). These findings do not seem to support Chen and Baker (2010) and Hyland (2008b) where the difference between the groups of writers in their studies was larger. For instance, in both studies, research articles included much more NP with of-phrase fragments than non-native student texts. On the other hand, the finding of the current study is consistent with Wei and Lei's (2011) indicating similar distribution of structures in non-native postgraduate texts and professional writing. Perhaps this is due to the fact that both the current study and Wei and Lei's study included texts from disciplines (i.e. foreign language teaching and applied linguistics, respectively) that require a high level of English even at undergraduate level. Therefore, the writers of these texts presumably have advanced English proficiency.

3.4. Functions of lexical bundles

The distribution of functional categories across the three groups of writers is represented in Figure 2. As can be seen, native postgraduate students and scholars used more referential bundles to make reference to entities, either physical or abstract, or to the textual context itself (e.g. *in the current study, at the end of, can be seen in, the total number of*).

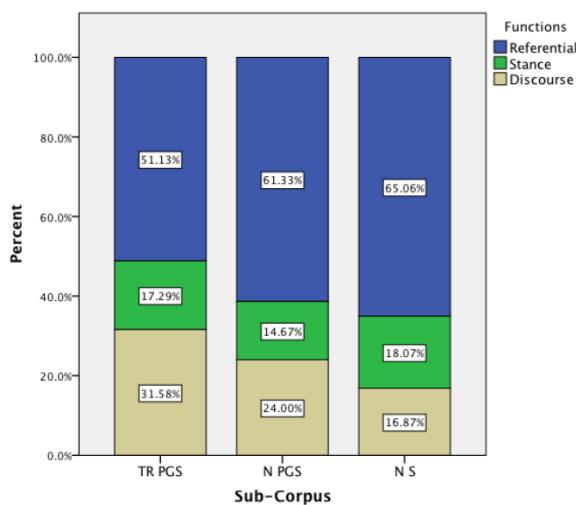


Figure 2. Functional distribution of lexical bundles (types)

Higher use of referential bundles by native writers was also found in Chen and Baker (2010) and although very slightly in Adel and Erman (2012). Leaving out the size of difference, the proportions of discourse and stance bundles are also similar. In these two studies, native texts included more stance

bundles to express writer attitude or assessment of certainty (e.g. *it is possible that, it may be that*) which was also the case in the current study.

Although Chen and Baker (2010) did find a significant difference, the chi-square test in the current study did not reveal any statistically significant differences between the distribution of functional categories and three groups of writers, $X^2(4, N = 291) = 6.67, p = .15$. The reason that Chen and Baker (2010) found a significant difference and this study did not could be attributed to the previously mentioned characteristic of the research corpus used in this study: the non-native students (i.e. Turkish postgraduates) were actually theses writers with advanced level English, not essay writers at undergraduate level like in their study. Therefore, this could explain why the chi-square test did not find a significant difference. Lastly, it should be noted that although there was no significant difference in the distribution of functions, the Turkish postgraduates employed different bundles, particularly stance bundles, in the same discourse function.

4. Conclusions

This study aimed to examine the use of lexical bundles in a corpus of MA and PhD theses produced by Turkish and native English postgraduate students, and published research articles by native English scholars in the area of foreign language teaching research. As a result of the analysis, a total of 271 4-word combinations occurring at 25 per million words and appearing in at least 5 different texts were identified in the research corpus. The highest number of bundle types was found in Turkish students' texts including 125 bundles while native students' texts contained 69 bundles and scholars' 77 bundles. Although it was reported in the previous literature that non-native writers would produce fewer bundles overall (Erman, 2009; Howarth, 1998) and less varied ones (Granger, 1998; Lewis, 2009) than native writers, the current study revealed a different finding in this respect. The Turkish postgraduate students in the research corpus was observed to employ a much wider range of lexical bundle types than the native students and scholars, which is consistent with Hyland's (2008b) and Wei and Lei's (2011) studies. This consistence is argued to be due to the fact that both studies and the current study contained postgraduate theses and dissertations in the research corpora. On the other hand, studies such as Chen and Baker's (2010) and Adel and Erman's (2012) focusing on university-level argumentative essays supported the aforementioned hypothesis. Therefore, it can be concluded that variety in lexical bundle use may be affected by writing expertise since these writers employed a wider range of bundles while constructing their texts compared to their native peers. Moreover, considering the 50 most frequent lexical bundles, almost half of the bundles in Turkish students' texts were either similar to or variants of those found in native students' and scholars' text, which can be interpreted as Turkish postgraduate students being familiar to the bundles used by their native peers and scholars to a certain extent.

Although there were similar bundles shared by three groups of writers, Turkish students extremely overused most of these bundles when compared to native students and scholars. This finding with regard to redundancy in non-native texts is also supported by Chen and Baker (2010) and Hyland (2008b). It can be inferred that despite being familiar with the frequently used bundles, Turkish postgraduate students use more varied bundles than native English students and scholars in a way more repetitive nature.

In terms of the significant differences in the frequency of actual bundles types, the current study revealed key findings. Firstly, 42 bundles were found to be statistically significantly overused by Turkish postgraduate students and 27 of these such as *it can be said that* and *it was seen that* were the bundles not shared with native English postgraduate students and scholars and argued to be unique to Turkish students. A comparison of the lexical bundles in similar studies showed that 11 of the 27 bundles overused by Turkish students but rarely or never used by native students and scholars were

not employed by non-native writers of different L1, either. This finding could be explained by some expressions in Turkish academic writing being transferred to English by the Turkish postgraduate students. For example, *it can be said + (that)* that was used by the Turkish students seems to be the English equal for one of the 10 most frequent verbs in academic Turkish, *söylenebilir*. Secondly, when compared to Turkish postgraduate students, native postgraduate students statistically significantly overused a total of 32 bundles and underused 9 bundles; as for native scholars, they overused 46 bundles, and underused 7 bundles. In other words, the current study revealed lexical bundles unique to Turkish postgraduate students and those unique to native postgraduate students and scholars. As a result, it can be concluded that in their use of lexical bundles while structuring their texts, Turkish postgraduate students, to a large extent, differed from their native peers and scholars in the area of foreign language teaching research.

As for the structural and functional analysis, the current study did not reveal any statistically significant differences between the three groups of writers included in the research corpus. There are only slight differences in the distribution of lexical bundles in both structural and functional categories, but these were also observed in Wei and Lei (2011). This finding may be due to the Turkish students' presumably high level of English owing to their area of study, i.e. foreign language teaching research. However, the extreme repetitive nature in the Turkish students' text was also observed here. Moreover, in spite of employing similar percentages of functions, they employed different bundles, especially stance bundles. Therefore, it can be deduced that Turkish postgraduate students employ similar proportions of structures and functions, but they make redundant use of bundles and employ different bundles although they seem to be using lexical bundles functionally and structurally at similar proportions.

Although Biber et al. (1999) argues that lexical bundles are very common and easily acquired in the natural discourse of language learning, Turkish postgraduate students whose MA and PhD theses were included in the research corpus seem not to have mastered the use of certain lexical bundles employed by native English postgraduates and scholars. According to Cortes (2004), this difference might be due to the lack of formal instruction given to the students in different disciplines on the frequency and function of such expressions. Regarding formal instruction, Eriksson (2012) suggested that while presenting lexical bundles in class, disciplinarity and specialization need to be considered when deciding what bundles to include. In this sense, the bundles identified to be commonly used by native students and scholars in the current discipline-specific study can be incorporated in academic writing courses of ELT programs. Similarly, those bundles found to be used by only Turkish postgraduate students can also be integrated in these courses in a way to make students notice that they can sometimes produce such bundles which may not seem native-like or academic. As discussed above, Turkish postgraduate students also made redundant use of certain bundles. Incorporating the key bundles reported in studies such as the current study in academic writing classes can enhance students' repertoire of lexical bundles, which may decrease the level of redundancy in their use of lexical bundles.

Several practices can be seen in Cortes (2006) and Eriksson (2012) on how such bundles can be incorporated in teaching. In this regard, functionally related lexical bundles taken from texts in a specific discipline can be introduced to students in contextualized examples. Students can be asked to analyse the functions and possible uses of these bundles. This can be followed by some application exercises including filling in the blanks, multiple choice or inappropriate use correction (Cortes, 2006). Different from these, students can be asked for their beliefs about usage of lexical bundles. For instance, they can be asked to choose which lexical bundle they think is commonly used in their discipline for a specific function. They can then be asked to use lexical bundles in the context of their own writing (Eriksson, 2012).

Further studies can investigate lexical bundles in different disciplines so as to guide student writers in their writing processes. Furthermore, the bundles unique to non-native writers or students with the same L1, as revealed in this study, can be investigated elaborately to identify whether it is simply transfer from L1. In addition to using a corpus including texts only in English, a parallel corpus in Turkish can also be combined in a further research, which may explain possible unique uses of Turkish writers in English can be attributed to the nature of Turkish in terms of commonly used words or expressions. A final suggestion would be on including non-contiguous word combinations along with contiguous word combinations such as lexical bundles in corpus-based studies. For instance, *play a role in* was identified as a four-word lexical bundles in the current study, but since non-contiguous combinations was not our focus, we did not discuss variations such as *play a vital/important/crucial role in*. Since the study of non-contiguous word combinations does not ignore variations within clusters maximizing the uncovering of word associations, it has been very popular in the last few years. Such combinations can also have great pedagogical value as they can serve frames for student writers.

References

Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81–92.

Bal, B. (2010). Analysis of Four-word Lexical Bundles in Published Research Articles Written by Turkish Scholars. Unpublished MA Thesis. Georgia State University, U.S.A.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263–286.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at: Lexical Bundles in University Teaching and Textbooks (A. Wilson, P. Rayson, & T. McEnery, Eds.). *Applied Linguistics*, 25(3), 371–405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Pearson Education Limited.

Biber, D., Conrad, S. Reppen, P. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Björkman, B. (2013). *English as an Academic Lingua Franca: An Investigation of Form and Communicative Effectiveness*. Berlin/New York: De Gruyter Mouton.

Byrd, P. & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, 5, 31-64.

Chen, Y. -H. (2009). Lexical Bundles across Learner Writing Development. Unpublished doctoral thesis. Lancaster University, Lancaster, UK.

Chen, Y., & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning & Technology*, 14(2), 30–49.

Cortes, V. (2002). Lexical bundles in Freshman composition. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds.), *Using corpora to explore linguistic variation* (pp. 131–145). Amsterdam: John Benjamins Publishing Company.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 397–423.

Cortes, V. 2006. Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education* 17: 391-406.

Cowie, A. (1998). Introduction. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 1-20). Oxford: Oxford University Press.

De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. *Learner English on computer* (pp. 67-79). London: Longman.

Erman, B. (2009). Formulaic language from a learner perspective: What the learner needs to know. In B. Corrigan, H. Quali, E. Moravcsik, & K. Wheatley (Eds.), *Formulaic language Vol. 2* (pp. 323-346). Amsterdam: John Benjamins.

Eriksson, A. (2012). Pedagogical perspectives on bundles: Teaching bundles to doctoral students of biochemistry. In James Thomas & Alex Boulton (eds). *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, 195-211.

Flowerdew, L. (2002). The exploitation of small learner corpora in EAP materials. In M. Ghandessy, A. Henry and R. L. Roseberry (Eds.) *Small Corpus Studies and ELT: Theory and Practice*. Amsterdam: John Benjamins.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145-160). Oxford: Clarendon Press.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24-44.

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32, 150-169.

Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.

Hyland, K. (2008b). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62.

Karabacak, E. & Qin, J. (2012). Comparison of lexical bundles by Turkish, Chinese, and American university students. Paper presented at Akdeniz Language Studies Conference at Akdeniz University. Antalya: Turkey.

Koester, A. 2010. Building small specialised corpora. In McCarthy, M. and O'Keeffe, A. (eds), *The Routledge Handbook of Corpus Linguistics*. London: Routledge.

Lewis, M. (2009). The idiom principle in L2 English: Assessing elusive formulaic sequences as indicators of idiomaticity, fluency, and proficiency. Saarbrücken, Germany: VDM Verlag.

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31(1), 25-35.

Scott, M. (2011). WordSmith Tools version 6. Liverpool: Lexical Analysis Software.

Simpson-Vlach, R., & Ellis, N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31, 487-512.

Sinclair, J. (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Tribble, C. (2002). Small corpora and teaching writing. In M. Ghandessy, A. Henry and R. L. Roseberry (Eds.) *Small Corpus Studies and ELT: Theory and Practice*. Amsterdam: John Benjamins.

Wei, Y. & Lei, L. (2011). Lexical Bundles in the Academic Writing of Advanced Chinese EFL Learners. *RELC Journal*, 42(2), 155-166.

Yıldız, İ. & Aksan, M. (2013). Türkçe bilimsel metinlerde eylemler: Derlem temelli bir inceleme. 27. Ulusal Dilbilim Kurultayı, 2-4 Mayıs 2013, Antalya, Kemer.

Zamel, V (1998). Questioning Academic Discourse. In V. Zamel and R. Spack (Eds.), *Negotiating academic literacies: Teaching and learning across languages and cultures* (pp. 187-197). Mahwah, NJ: Erlbaum.

Anadili Türkçe ve İngilizce Olan Yazarların Sözcük Öbeği Kullanımı

Öz

Diger yandan (on the other hand) ve sonuç olarak (as a result of) örneklerindeki gibi İngilizce'deki sözcük öbekleri oldukça yaygın olarak kullanılır ve akademik söylemde önemlidir. Belirli bir disiplinde tipik olarak kullanılan sözcük öbeklerinin başarılı bir şekilde kullanılması yazarlar için önem taşıır ve bu öbeklerin olmaması okuyucuya akıcı ve anadil konuşuru gibi gelmez. Son zamanlardaki çalışmalar (örn. Adel & Erman, 2012; Chen & Baker, 2010) anadili İngilizce olmayan yazarların daha az sayıda ve çeşitte sözcük öbeği kullandıklarını göstermiştir. Bu konuya odaklanan bu çalışmada anadili Türkçe ve İngilizce olan lisansüstü öğrencilerin ve anadili İngilizce olan akademisyenlerin sözcük öbeği kullanımlarının sıklık, işlev ve yapı açısından incelemesi amaçlanmıştır. WordSmith Tools 6 yazılımı kullanılarak dört-sözcüklü sözcük öbekleri belirlenmiştir. Sonuçlar, Türk lisansüstü öğrencilerin anadil konuşuru öğrenci ve akademisyenlere oranla çok daha fazla sayıda sözcük öbeği kullandıklarını göstermiştir. Ancak, sıklıklar incelendiğinde Türk öğrencilerin metinlerinde gereksiz düzeyde bir aşırı kullanım olduğu görülmüştür. Diğer yandan, öbek listelerinin istatistiksel incelemesi göstermiştir ki Türk öğrenciler anadil konuşuru öğrenciler ve akademisyenlerden farklı öbekleri kullanmışlardır. Son olarak, kullanılan sözcük öbeklerinin yapısal ve işlevsel kategorileri araştırma alt-derlemleri arasında istatistiksel olarak herhangi bir anlamlı fark göstermemiştir.

Anahtar sözcükler: Sözcük öbekleri; İngilizce akademik söylem; Türk öğrenciler; İngilizce anadil konuşuru öğrenciler; derlem dilbilim

AUTHOR BIODATA

Yusuf ÖZTÜRK is a research assistant at the English Language Teaching Department of Anadolu University, Eskisehir, Turkey. He is currently a PhD student at the same department. His research interests include corpus linguistics, academic discourse, formulaic language and technology integration into teaching.

Gül DURMUŞOĞLU KÖSE works as a professor of English Language Teaching at Anadolu University Education Faculty in Eskisehir, Turkey. She had M.A and PhD degrees from Exeter University in 1980 and 1984. She has been giving lectures on applied linguistics, translation, sociolinguistics, discourse and research skills at Anadolu University. She is currently the director of Foreign Languages Research Center at Anadolu University. She has many published research papers and she has supervised various theses on second language acquisition, language learning and teaching. Her research interest includes translation studies, discourse, pragmatics and sociolinguistics.